

TEMU KEMBALI CITRA DIGITAL BERDASARKAN FITUR WARANA DAN TEKSTUR DENGAN KOMBINASI FUNGSI KLUSTERING TERPISAH DAN PEMBOBOTAN MANUAL

Mirza Galih Kurniawan¹, Hamdan Anang Kholili²

Mahasiswa Pasca Sarjana Teknik Informatika ITS

e-mail : ¹⁾ mirza_mail_box@yahoo.co.id, ²⁾ hamdan.ak86@gmail.com

Abstrak. Temu kembali citra digital yang didasarkan pada fitur warna dan tekstur sudah cukup banyak dilakukan. Sebagian besar penggunaan dua fitur tersebut dijadikan satu vektor fitur yang selanjutnya vektor fitur yang didapat diukur kedekatannya dengan vektor fitur citra query. Dalam penelitian ini diterapkan pengelompokan (*clustering*) untuk masing-masing fitur. Selanjutnya pengukuran kedekatan dari vektor fitur citra query diukur dengan centroid tiap-tiap kluster. Kluster dari centroid tiap fitur yang memiliki jarak terdekat dengan vektor fitur citra query dijadikan kandidat vektor terpilih. Selanjutnya perhitungan pembobotan dan similaritas dengan citra query hanya dilakukan pada gabungan kluster terpilih dari tiap-tiap fitur dengan vektor fitur citra query. Sehingga proses pembobotan tidak dilakukan pada keseluruhan citra dalam dataset. Selain itu diterapkan juga penggabungan penilaian melalui klustering manual yang diharapkan dapat memberikan kontribusi hasil yang lebih bagus melalui pendekatan semantik.

Kata Kunci: CBIR, Color, Texture, Klustering, K-Means, SURF

Content Based Image Retrieval atau CBIR merupakan salah satu bentuk aplikasi *computer vision* untuk pencarian citra berdasarkan fitur yang ada pada citra itu sendiri.

Warna dan tekstur merupakan fitur penting yang dapat digunakan dalam CBIR [3] sehingga warna dan tekstur banyak digunakan sebagai fitur yang digunakan dalam membentuk vektor fitur sebuah citra pada proses CBIR.

Proses CBIR pada umumnya mengukur kedekatan vektor fitur dari citra *query* dengan vektor fitur pada database dengan berbagai algoritama. Chen [2] menggunakan kombinasi klustering vektor pada proses CBIR yang dikenal dengan metode CLUE. Swati [7] menggunakan fitur warna dan tekstur pada proses CBIR dengan pendekatan *wavelet* pada ekstraksi fiturnya.

Proses CBIR yang tidak menambahkan proses klustering secara otomatis mengukur keseluruhan vektor fitur pada database. sedangkan metode CBIR yang menggunakan *klustering* menggunakan satu atau beberapa gabungan fitur yang dijadikan satu proses klustering sehingga tidak dapat diprioritaskan untuk fitur yang memang dijadikan prioritas.

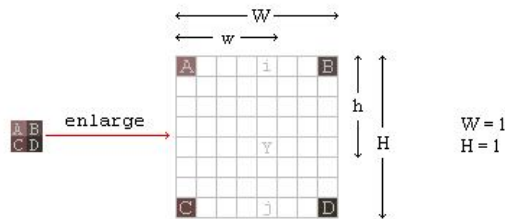
Pada penelitian ini diperkenalkan pendekatan klustering baru pada proses CBIR dengan menggunakan fungsi klustering terpisah. dengan pemisahan proses pembobotan melalui klustering terpisah diharapkan dapat memperbaiki kinerja dari sisi pembobotan otomatis yang didasarkan pada warna dan tekstur. Selain itu tiap fitur akan dapat disesuaikan presentase bobot yang diinginkan.

Penilaian manusia terhadap citra hasil query bukan hanya didasarkan pada kemiripan warna dan tekstur namun lebih cenderung ke makna dari suatu citra atau dalam hal ini adalah nilai semantik dari citra. Untuk itu digabungkan metode *clustering* manual untuk dapat mengakomodir pendekatan CBIR melalui semantik. Klustering manual ini merupakan hasil dari penelitian yang dilakukan oleh fauzy [7].

BILINEAR INTERPOLATION

Bilinear interpolation merupakan salah satu metode *resampling* dimana *resampling* merupakan metode matematis untuk membuat sebuah versi baru dari citra dengan ukuran yang berbeda dari citra aslinya. Perhitungan piksel

baru pada *bilinear Interpolation* terdapat pada persamaan 1



$$Y = A + w(B - A) + h(C + w(D - C) - (A + w(B - A)))$$

$$Y = A(1 - w)(1 - h) + B(w)(1 - h) + C(h)(1 - w) + D(wh) \quad (1)$$

SURF

Mencari titik korespondensi antara dua gambar objek yang sama merupakan bagian dari banyak aplikasi visi komputer. SURF (*Speed Up Robust Feature*) [5] merupakan salah satu detektor *interest point* lokal yang telah digunakan banyak aplikasi visi komputer saat ini. SURF telah terbukti mampu mencapai pengulangan yang tinggi dalam mengenali kekhasan suatu gambar.

Fitur SURF adalah *keypoint* dari sebuah gambar. *Keypoint* adalah titik-titik dari sebuah gambar yang nilainya tetap ketika mengalami perubahan skala, rotasi, *blurring*, pencahayaan, dan juga perubahan bentuk. Perubahan bentuk ini bisa terjadi karena bentuk gambar query yang tidak utuh atau tidak sempurna seperti gambar yang ada didalam database gambar tersebut. Gambar query yang tidak utuh mungkin karena objek lain yang menutupi, atau pengambilan gambar yang tidak sempurna, atau keadaan objek itu sendiri yang mengalami perubahan.

Algoritma SURF merupakan gabungan dari skala dan deteksi *interest point* rotasi-invarian. Metode ini menggunakan integral gambar untuk meningkatkan kecepatan dalam komputasi. Poin-poin penting dideteksi dengan menggunakan matriks *Fast-Hessian*. Integral gambar dihitung dari gambar input kemudian dihitung setiap area pada gambar tersebut.

K-Means

K-means merupakan metode *flat clustering* dimana metode klusterisasinya didasarkan pada minimalisasi rata-rata jarak euclidian suatu elemen terhadap pusat kluster nya. Pusat kluster tersebut sering diistilahkan dengan nama sentroid [6]. Hasil dari proses pengelompokkan

data menggunakan *k-means* sangat bergantung pada nilai awal sentroid yang diberikan. Nilai awal sentroid pada *k-means* diberikan secara random sehingga pada hasil klusterisasi menggunakan *k-means* sering terjebak lokal optimal dari nilai sentroid awal yang diberikan [1]. Pada penelitian ini *k-means* di gunakan untuk proses klustering otomatis yang mengelompokkan dataset ke dalam dua grup kluster yang berbeda.

Precision dan Recall

Jika kita mengasumsikan A sebagai himpunan dokumen yang seharusnya dikembalikan oleh sistem temu kembali (*retrieved document*), B adalah himpunan dokumen yang relevan dengan *query*, dan $A \cap B$ adalah himpunan dokumen benar yang diberikan sistem temu kembali informasi, maka nilai *recall* dan *precision* sistem temu kembali informasi tersebut dapat dinyatakan sebagai berikut :

Precision

$$= \frac{|A \cap B|}{|A|} \quad (2)$$

recall

$$= \frac{|A \cap B|}{|B|} \quad (3)$$

Pada dasarnya, nilai *recall* dan *precision* berada pada rentang antara 0 sampai dengan 1 atau dapat juga dinyatakan dalam prosentase. Oleh karena itu, suatu sistem temu kembali yang baik adalah yang dapat memberikan nilai *recall* dan *precision* mendekati 1 atau 100% . dalam penelitian ini hanya digunakan *precision* untuk mengevaluasi hasil yang didapat.

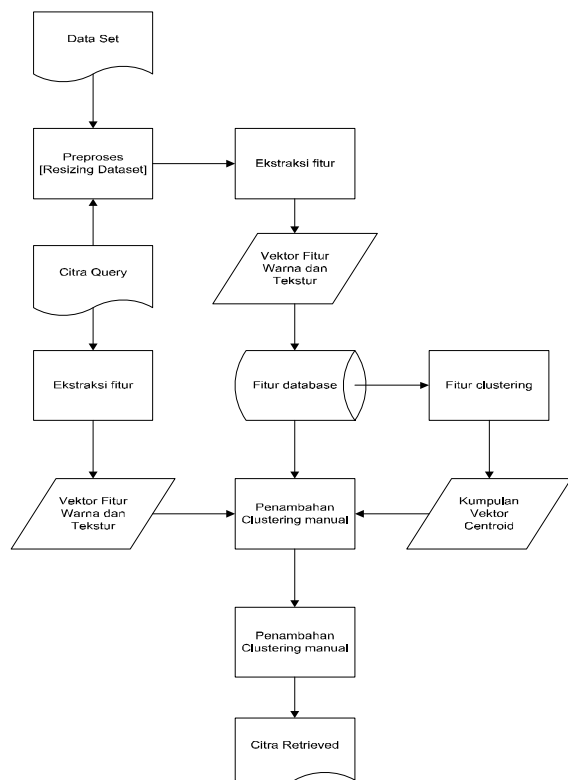
METODOLOGI

Secara garis besar. Metodologi yang dipakai pada penelitian ini seperti yang terdapat pada gambar 1. Untuk detail masing masing tahapan terdapat pada sub bab selanjutnya

Dataset

Dataset yang dipakai dalam penelitian ini merupakan dataset kumpulan citra. Citra tersebut didapat dari beberapa *website* yang menyediakan kumpulan gambar yang sudah dikategorisasikan oleh *website* tersebut. Gambar dari beberapa *website* tersebut di simpan sesuai dengan kategrosi masing-masing.

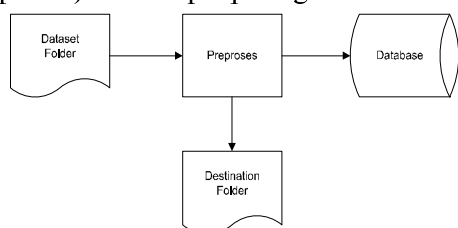
Setiap kategori tersebut mewakili proses *clustering* secara manual.



Gambar 1. Garis Besar Algoritma yang Digunakan

Preproses

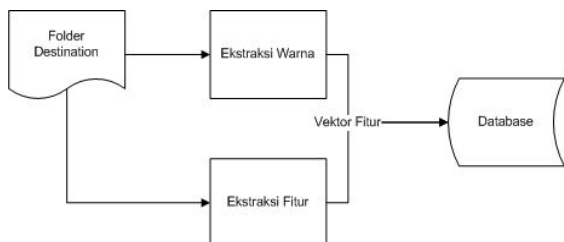
Dalam tahapan preproses ini dilakukan proses *resizing* citra dengan metode *bilinear interpolation*. Setiap citra yang sudah di-*resize* dimasukkan ke dalam database sesuai dengan *path file* dan kategorinya masing-masing. Tujuan dari *resizing* citra ini adalah untuk mempercepat proses ekstraksi fitur warna dan fitur tekstur pada gambar yang memiliki dimensi (ukuran) yang besar. Hasil dari proses *resizing* merupakan *file* citra hasil *resizing* yang disimpan pada folder tersendiri dan *record* dari dataset yang disimpan pada database. Selanjutnya *file – file* hasil *resizing* tersebut yang digunakan pada proses – proses selanjutnya. Diagram proses *resizing* (preproses) ini terdapat pada gambar 2.



Gambar 2. Diagram Alur Proses pada Tahap Preproses

Ekstraksi Fitur

Tahapan setelah preproses adalah ekstraksi fitur dari citra hasil preproses. Fitur yang dipakai dalam penelitian ini adalah fitur warna (dalam RGB) dan tekstur/karakteristik citra dengan menggunakan metode SURF. Hasil dari ekstraksi fitur adalah vektor fitur dari citra hasil preproses yang selanjutnya vektor fitur tersebut disimpan pada database. Diagram proses pada proses ekstraksi fitur terdapat pada gambar 3.



Gambar 3. Digram Proses Ekstraksi Fitur

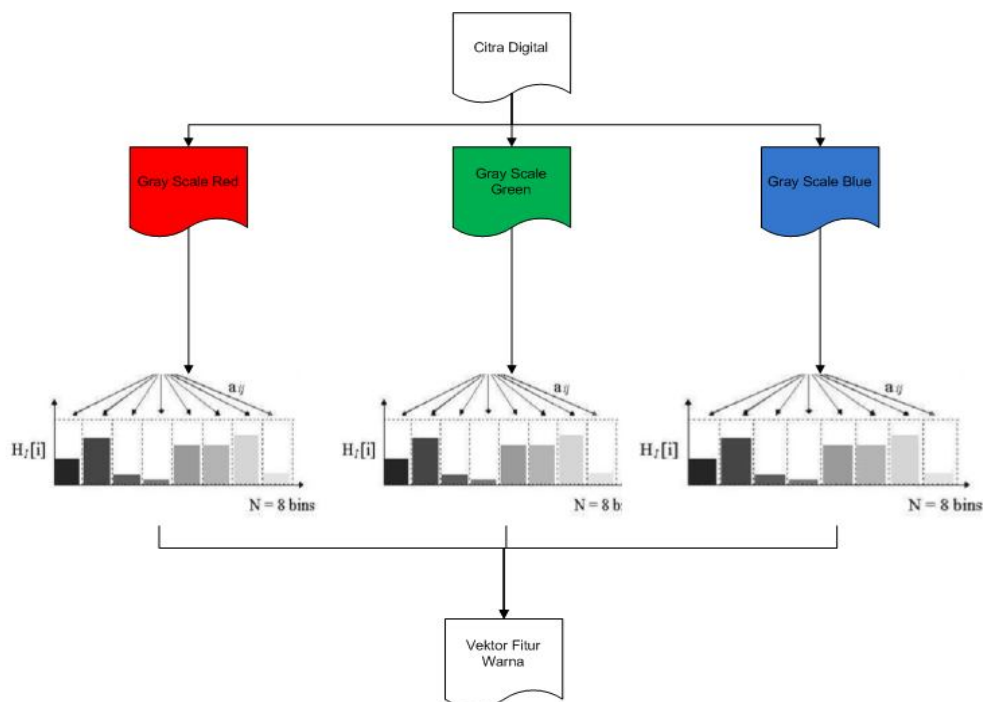
Ekstraksi Fitur Warna

Dalam proses ekstraksi fitur warna digunakan vektor fitur berupa nilai frekuensi warna tiap selang warna tertentu. Dalam penelitian ini digunakan warna RGB dan tiap *channel* warna di buat histogram frekuensinya masing masing dibagi kedalam 8 *bin*. Gabungan dari nilai frekuensi tiap *bin* dalam RGB disimpan didalam vektor fitur warna di dalam database. Panjang dari vektor fitur warna yang digunakan adalah 3x8 elemen. Nilai tiap elemen dalam vector merupakan nilai frekuensi piksel dalam range yang merupakan hasil pembagian dari nilai 256 dibagi ke dalam 8 bin untuk setiap *channel* warna ($R|G|B$).

Ekstraksi Fitur Tekstur dengan Metode Surf

Implementasi proses ekstraksi menggunakan surf dilakukan dengan menggunakan Emgu CV dengan menggunakan point interest sebanyak 20. Dari proses ekstraksi didapat 80 buah elemen vektor point yang digunakan sebagai fitur yang akan disimpan di database.

Proses yang pertama kali dilakukan dalam metode surf adalah membuat ruang skala (*scale space*). Proses ini dilakukan agar invarian terhadap perubahan skala pada citra. *Scale space* terbagi kedalam bilangan yang disebut *octave*. Setiap *octave* merepresentasikan respon filter yang diperoleh dengan melakukan proses

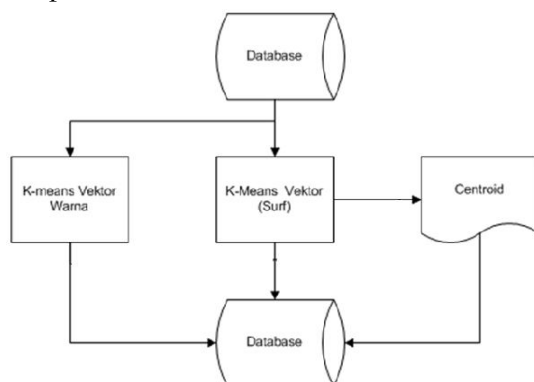


Gambar 4. Diagram Proses Ekstraksi Fitur Warna

konvolusi gambar yang diinputkan. Proses selanjutnya adalah lokalisasi *keypoint*, yaitu proses mencari lokasi *keypoint* menggunakan interpolasi data yang dekat dengan *keypoint* hasil proses sebelumnya.

Klustering Otomatis

Dari masing-masing fitur yang sudah diekstraksi (fitur warna dan fitur tekstur) akan dilakukan klusterisasi. Proses klusterisasi ini menggunakan k-means pada masing-masing fitur. Sehingga terdapat dua proses *clustering* yaitu pada fitur warna dan fitur tekstur. Hasil akhir pada tahapan ini adalah adanya beberapa kluster pada masing-masing fitur. Setiap sentroid dari setiap kluster tersebut akan di simpan di dalam database.



Gambar 5. Diagram Proses Klustering Otomatis

Klustering Manual

Untuk mewakili *expert* (Ahli) dalam penentuan klustering, digunakan metode pengkategorian gambar secara manual yang di ambil dari beberapa website yaitu <http://popularscreensavers.com>, <http://webtaj.com> dan <http://www.9freepictures.com>. Dari setiap website tersebut didapatkan dataset gambar yang telah terbagi dalam beberapa kategori. Dalam penelitian ini kategori yang akan digunakan adalah *Sports* (571 gambar), *flower* (581 gambar) dan *animals* (841 gambar). Kategori gambar tersebut akan digabungkan dengan hasil *clustering* citra fitur dan warna. Untuk data uji dilakukan dengan mengambil data gambar dari website <http://hdwallpaperspick.com> secara acak dengan kategori gambar yang sama dengan dataset yang di ambil.

Proses klustering manual didasarkan pada pengelompokan pada posisi kategori gambar pada sumber dataset. Selanjutnya file gambar yang memiliki kategori yang bersesuaian dijadikan satu direktori pada dataset yang akan digunakan. Satu direktori tersebut mewakili satu kluster manual. Kandidat kategori yang akan digunakan adalah kategori yang bersesuaian dengan file citra pada dataset yang memiliki bobot total fitur warna dan kluster yang terbesar.

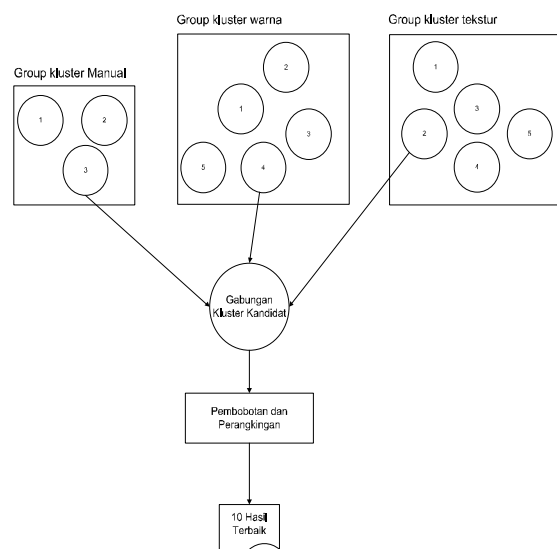
Kandidat yang akan dihitung prioritas dan bobot akhir adalah gabungan dari kandidat dari fitur warna, tekstur dan klustering manual.

Pembobotan

Pada tahapan pembobotan semua kluster dari dataset sudah terbentuk. Terdapat dua grup kluster yang terdiri dari grup kluster fitur warna dan fitur tekstur. Proses *retrieve* citra dilakukan dengan terlebih dahulu membandingkan vektor fitur dari citra query dengan masing-masing centroid dari kluster. Kluster yang memiliki *centroid* terdekat dengan citra *query* dijadikan kluster kandidat. Kluster kandidat berikutnya adalah hasil dari *clustering* manual.

Setiap vektor fitur dalam kluster kandidat tersebut digabung dalam satu kumpulan vektor kandidat. yang selanjutnya fitur kandidat tersebut diukur similaritasnya dengan citra query.

Hasil akhir yang didapat adalah urutan vektor fitur dari kumpulan vektor kandidat yang mempunyai jarak terdekat dengan vektor fitur citra query.



Gambar 6. Penggabungan Kluster Kandidat dan Pembobotan

Evaluasi

Pengujian dilakukan dengan melakukan *query* pada *dataset* yang ada dengan citra uji yang sesuai dengan kategori tertentu pada website acuan.

Hasil validitas hasil *query* merupakan kesesuaian antara kategori pada website yang menjadi dataset dengan hasil *query* dari proses CBIR. Pada pengujian *query* tersebut akan di

ambil 10 data sebagai data hasil *retrieve*. Dari hasil uji coba tersebut akan dihitung nilai *precision* pada setiap *query* yang dilakukan.

Tabel 1. Hasil Evaluasi

No	Citra Uji	Precision (%)
1	Earth-Flower.jpg	100
2	Black-Flower.jpg	90
3	Alessandro_Del_Piero.jpeg	100
4	cristiano-ronaldo.jpg	100
5	rock-climbing.jpg	100
6	CitizensBankPark.jpg	90
7	extreme-sports-surfing-12-6.jpg	100
8	african-elephants-8-9.jpg	100
9	african-lions-77-9.jpg	100
10	african-zebra-9-21.jpg	90
11	albino-animals-from-snowflake-the-white-gorilla-diamond.jpg	90
12	baby-pig-68-7.jpg	100
13	beautiful-white-swan.jpg	100
14	chihuahua-dog.jpg	100
15	cute-kitten-83-4.jpg	100
16	beautiful-red-rose-flower.jpg	90
17	beautiful-orchid-22-27.jpg	100
18	beautiful-rainbow-rose-flower.jpg	80
19	beautiful-tulip-24-9.jpg	100
20	liverpool-logo-gold.jpg	90
21	grizzly-bear-82-5.jpg	100
22	happy-animal.jpg	90
23	squirell_monkey-150x150.jpg	90
24	Love-Rose-HD-Wallpapers-150x150.jpg	100
25	2010_fifa_world_cup_footb_all_stadium-t2.jpg	100
26	baseball.jpg	100
27	Flowers-HD-Wallpaper-2560x1440.jpg	90
28	free-green-rose-desktop-wallpaper.jpg	100
29	Beautiful-Colorful-Cute-Birds-Wallpapers.jpeg	100
30	funny-animal-backgrounds.jpg	100

Tabel di atas menunjukkan hasil percobaan dengan melakukan 30 kali query. Dari tabel hasil percobaan tersebut telah menunjukkan nilai *precision* yang tinggi dari seluruh hasil percobaan. Dari 10 data yang berhasil di retrieve, terdapat data antara 8 sampai 10 data yang relevan sehingga menghasilkan nilai *precision* yang tinggi yaitu antara 80% hingga 100%. Dari data hasil uji coba diatas menunjukkan bahwa dengan menggabungkan fitur warna, tekstur dan pengkategorian secara manual dalam query citra telah memberikan hasil yang sangat baik dengan nilai *precision* rata-rata mencapai 96% dalam 30 kali percobaan yang telah dilakukan. Dalam penelitian ini digunakan nilai *precision* dikarenakan adanya batasan dalam jumlah citra ter-*retrieve*.

SIMPULAN

Pada penelitian ini digunakan fungsi ekstraksi dan *klustering* sederhana karena pada penelitian ini dimaksudkan untuk memperkenalkan pendekatan pada kombinasi fungsi *klustering* terpisah dan pembobotan manual. Untuk pengembangan pada metode dalam ekstraksi dan *klustering* dapat dijadikan bahan dalam penelitian – penelitian selanjutnya

PUSTAKA

- [1] Barakbah, Ali Ridho. 2011. *A Pillar Algorithm for K-Means Optimization by Distance Maximization for Initial Centroid Designation*. IEEE
- [2] Chen, Yixin . 2003. Content-Based Image Retrieval by Klustering. ACM
- [3] Kahitna, Ch. 2011. Image Retrieval Based On Color and Texture Features of the Image Sub-blocks . International Journal of Computer Applications (0975 – 8887) Volume 15– No.7, February 2011
- [4] Fauzi, Ali .2013. Term Weighting Berbasis Indeks Buku Dan Kelas Untuk Perangkingan Dokumen Berbahasa Arab. ITS
- [5] Herbert Bay. SURF: Speeded Up Robust Features. Computer Vision and Image Understanding (CVIU)
- [6] Manning. 2008. Christopher D. et all. *Introduction to Information Retrieval*. Cambridge University Press.
- [7] Swati, V Sakhare. 2011. Design of Feature Extraction in Content Based Image

Retrieval (CBIR) using Color and Texture. International Journal of Computer Science & Informatics, Volume-I, Issue-II